Vol. 184, No. 3, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 15, 1992 Pages 1468-1476

STRUCTURE OF HUMAN PROSTATIC ACID PHOSPHATASE GENE

Farida S. Sharief and Steven S.-L. Li*

Laboratory of Genetics
National Institute of Environmental Health Sciences
National Institutes of Health
Research Triangle Park, North Carolina 27709

Received April 2, 1992

Two cDNA clones containing the complete protein-coding sequence of 1,188 nucleotides as well as the 5' and 3' non-coding regions of human prostatic acid phosphatase (PAP) were isolated and sequenced. The size of PAP mRNAs from benign prostate hyperplasia and cancerous prostate was estimated to be 3.2Kb, indicating that the 3' downstream polyadenylation signal was used. Several genomic clones containing parts of the human PAP gene were isolated and the nucleotide sequence of ten exons and their flanking regions was determined. The protein-coding sequence of the human PAP gene was interrupted by nine introns. The positions of all nine introns present in the human PAP gene were homologous to those of the first nine introns in the human lysosomal acid phosphatase (LAP) gene. However, the last (11th) exon of the LAP gene encoding the COOH-terminal domain, which includes a transmembrane segment, was found to be absent in human PAP gene. Southern blot analysis of ten mammalian genomic DNAs gave multiple EcoRI fragements. The data of human genomic DNAs were consistent with the total length of the PAP gene of at least 50 kilobases.

Acid phosphatases (orthophosphoric monoester phosphohydrolase, EC 3.1.3.2) are a group of enzymes capable of hydrolysing phosphomonoesters under acidic conditions. They can be differentiated according to their immunological properties, tissue distribution and subcellular location (1). Prostatic acid phosphatase (PAP) is synthesized under

* To whom all correspondence should be addressed.

Abbreviations: PAP, prostatic acid phosphatase; LAP, lysosomal acid phosphatase.

0006-291X/92 $4.00

androgen regulation by epithelial cells of the prostate and secreted into seminal fluid. Human PAP has been used for over 50 years as a tumor marker to diagnose patients and prognose with prostate cancer (2-3). The physiological function of PAP is still not well understood, but recent reports suggest that PAP exhibits protein tyrosine phosphatase activity (4-5).

In order to study protein structure-function and the mechanism of androgen regulation, we have undertaken protein sequencing, cDNA cloning and isolation of the human PAP gene. We previously described the amino acid sequence and antigenic structure of the secretory PAP enzyme (6-7). We also cloned the partial PAP cDNA sequences and mapped the PAP gene to chromosome 3 (6). Here we report the isolation of a nearly full-length cDNA and characterization of exon-intron organization of the human PAP gene as well as the results of Northern and Southern blot analyses.

MATERIALS AND METHODS

Cloning of full-length cDNA:

In order to isolate the cDNA clones containing 5' non-coding and protein-coding regions, the EcoRI*-EcoRI and EcoRI-Hind III DNA fragments (Fig. 1) of the human PAP cDNA clone hP40 reported previously (6) were labelled with (α-32P)dCTP and used as a probe to screen a prostate lambda gt11 cDNA library (Clontech) using standard procedures (8). The DNAs purified from positive cDNA clones were partially characterized by restriction endonuclease mapping and Southern blotting. Their EcoRI DNA fragments were subcloned into M13mp18 phage and the nucleotide sequences of the inserted DNAs were determined by the dideoxy chain termination method with the sequencing protocol using deoxyadenosine 5'-(α-35S) thiotriphosphate (9).

Isolation of genomic clones:

Human genomic clones containing parts of the PAP gene were isolated from a total lymphocyte genomic DNA library (Stratagene) using the cDNA probes of clone hP40 (6,8). Genomic clones containing the exon 1 and its upstream sequence of the PAP gene were subsequently isolated using cDNA probes of 5' coding-region from clone hP303 (see Fig. 1). The DNAs purified from positive genomic clones were characterized by restriction endonuclease mapping and Southern blotting. The DNA fragments exhibiting positive hybridization to the cDNA probe were subcloned into M13mp18/19 phage or Bluescript phagemid. The nucleotide sequences of the exons and their flanking regions were determined by dideoxy chain termination method using either universal or exon-specific primers (9).

Northern blot analysis:

Tissues of benign prostate hyperplasia and cancerous prostate from patients were kindly provided by Drs. Y. Sharief and J. Mohler, and were quickly frozen in liquid nitrogen, and total RNAs were extracted by the guanidinium thiocyanate-phenol-chloroform method using the RNAzol B kit (Biotecx Laboratories, Inc.). Poly(A)+RNAs were purified by oligo(dT) cellulose using the FastTrack kit (Invitrogen Corp.). Two micrograms of poly(A)+RNAs were electrophoresed on a 1.0% agarose/2.1% formaldehyde gel, transferred to a Nytran membrane (Schleicher & Schull) by capillary blotting with 10X SSC (1.5M sodium chloride/0.15M sodium citrate, pH7), and UV-cross-linked as described (8). The Northern blot was hydridized overnight at 42 C in 50% Formamide/ 5X Denhard's solution/ 5X SSPE (0.75M sodium chloride, 50mM sodium phosphate, 5mM EDTA, pH7.4)/ 1% SDS/ salmon sperm DNA (200 ug/ml) with randomly 32P-labelled PAP cDNA probe containing no Alu-sequence. The filter was rinsed twice at room temperature for 15 min each and washed twice at 65 C for 15 min each with 1X SSPE/ 0.1% SDS. Autoradiograms were obtained by exposing the blot to Kodak XAR-5 film with intensifying screen at -70 C for various lengths of times.

Southern blot analysis:

The Southern blot containing EcoRI fragments of ten mammalian genomic DNAs was obtained from BIOS Corporation and hybridized using 32P-labelled PAP cDNA probe containing no Alu-sequence according to the procedure recommended by the supplier.

RESULTS

Nucleotide sequence of PAP cDNAs:

Two cDNA clones hP303 and hP317 were isolated and partially characterized (Fig. 1). The nucleotide sequence of both strands of the inserted DNA from clone hP303 was determined and found to consist of complete protein-coding sequence of 1,188 nucleotides as well as 5' (42 nucleotides) and 3' (1136 nucleotides) non-coding regions. This PAP cDNA sequence was consistent with that reported previously (6,12).

Characterization of PAP gene:

With the use of the cDNA probe of clone hP40, several genomic clones containing parts of exons 2 through 10 were isolated (Fig. 2) and the nucleotide sequences of these nine exons and their flanking regions were determined. The 5' cDNA region, including exon 1, of the newly isolated clone hP303 was subsequently used as a probe to rescreen the genomic DNA

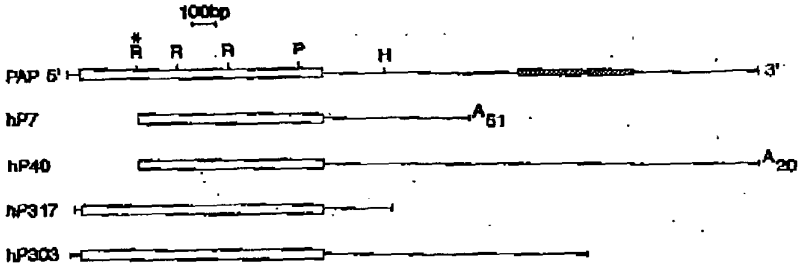

Fig. 1. Restriction endonuclease map of human PAP cDNA clones.

The restriction sites given (R, EcoRI; R*, EcoRI star activity; P, Pst I; H, Hind III) are those used in nucleotide sequencing and preparation of hybridization probes. The protein-coding sequence is denoted by an open box. The dimeric and monomeric Alu-repeats are shown by hatched boxes. Clones hP7 and hP40, which were reported previously (6), contain poly(A+)-tail of 51 and 20 A's, respectively, and their 5' sequences were presumably truncated by an EcoRI star activity during construction of the cDNA library. Clones hP303 and hP317 were isolated in this investigation.

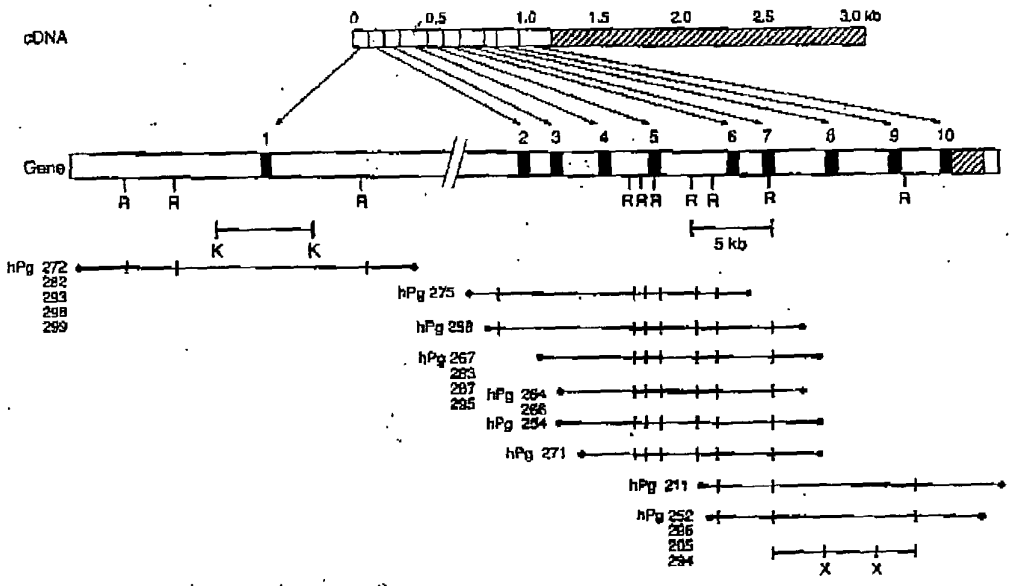

Fig. 2. The exon-intron organization of the human PAP gene along with restriction endonuclease sites.

The protein-coding and 3' noncoding regions of the PAP cDNA are indicated by open and hatched boxes, respectively. The exons in the PAP gene are numbered in Arabic and shown by solid-blocks. The exact sizes of nine introns are not known, and the intron between exon 1 and 2 was not overlapped by these isolated genomic clones. Only those restriction sites (R, EcoRI; K, KpnI; X, XbaI) which were used to isolate DNA fragments and to subclone are given.

library, and five identical clones were found to contain exon 1 and its 5' upstream region (Fig. 2). The nucleotide sequence of ten exons of the PAP gene was identical to the PAP cDNA sequence (Fig. 3).

The protein-coding sequence of the PAP gene was interrupted by nine introns. Exon 1 encodes a leader peptide

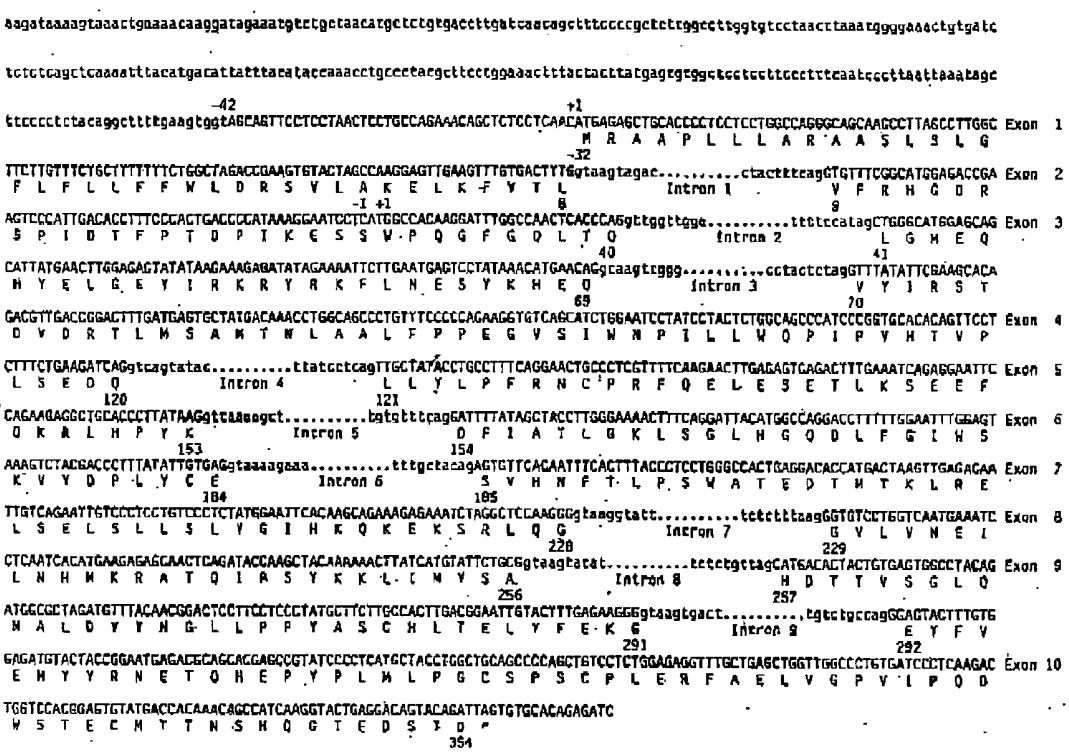

Fig. 3. Nucleotide sequences of the protein-coding exons and their flanking regions of the human PAP gene.

The nucleotide sequences of ten exons as well as cDNA are given in upper case, while those of flanking sequences are in lower case. The 5' non-coding region of clone hP303 contains 42 nucleotides, and the exact site of transcription initiation remains to be determined. The sizes of exons 2 through 9 are 96, 90, 153, 99, 93, 132, 96, and 104 nucleotides, respectively. The 3' non-coding region of 1,913 nucleotides present in cDNA sequence was reported previously (6), and only partial sequence determined in genomic DNA is presented. The deduced amino acid sequence is shown below the coding-exons, and is numbered from the first residue of the secretory enzyme. The 32 amino acids in the leader peptide are given in negative numbers. The translation termination codon TAG is denoted by *. It should be noted that a Cys instead of the previously reported Val (6) was identified at amino acid 340 of the secretory enzyme.

of 32 amino acids and the first eight amino acids of the secretory enzyme as well as a 5' non-coding region of at least 42 nucleotides. Exons 2 through 9 range from 90 to 153 nucleotides in size. The last (10th) exon encodes the COOH-terminus of 64 amino acids and 3' non-coding region of 649/1916 nucleotides, depending on the polyadenylation site. All exon/intron junction sites were consistent with the conserved gt/ag sequence, except gc/ag for intron 3 . The first eight introns of the PAP gene are class-0 introns located between codons. The ninth intron is a class-II intron located between 2nd and 3rd bases of codon 291. The length of these nine introns was not precisely determined, and the size of intron 1 was relatively large and not overlapped by these isolated genomic clones. The total length of the human PAP gene was estimated to be at least 50 kilobases.

Expression of PAP poly(A)+RNAs:

The results of Northern blot analysis of poly(A)+RNAs isolated from benign prostate hyperplasia and canerous prostate are presented in Fig. 4. The size of the most predominant PAP mRNAs was estimated to be 3.2Kb, although weak singal of smaller RNAs was detected in the specimen of cancerous prostate.

Southern blot analysis of mammalian genomes:

The EcoRI fragments of mammalian genomic DNAs were analyzed using PAP cDNA probe contianing no Alu-sequence (Fig. 5). Human genomic DNAs exhibited multiple EcoRI fragments, and most of the other mammalian DNAs gave three hybridization bands.

DISCUSSION

In our previous report (6), all PAP cDNAs were truncated at 5' protein-coding region because of the presence of EcoRI star activity during the construction of cDNA library. In this investigation, two newly isolated clones hP317 and hP303 contain 5' noncoding (42 nucleotides) and complete protein-coding (1136 nucleotides) sequences in addition to partial 3' noncoding region. The reported heterogeneous sizes (646 and

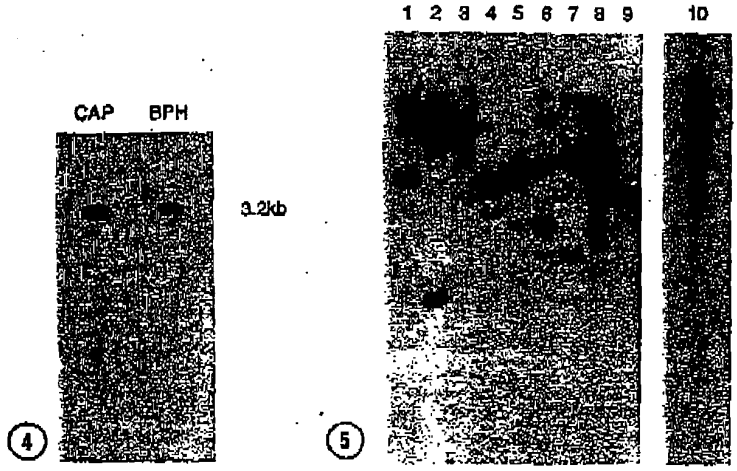

Fig. 4. Northern blot analysis of poly(A)+RNAs from human prostate tissues.
Tissues: CAP, cancerous prostate; BPH, benign prostate hyperplasia. The Northern blot was probed with PAP cDNA containing no Alu-sequence as well as human GAPDH cDNA. The size of GAPDH mRNAs was estiamted to be 1.3Kb as predicted (data not shown).

Fig. 5. Southern blot analysis of mammalian genomic DNAs.
Lanes 1, dog; 2, cat; 3, rabbit; 4, cow; 5, sheep; 6, mouse; 7, rat; 8, hamster; 9, pig; and 10, human. The autoradiograph of lanes 1 through 9 was exposed for 4 days, while lane 10 was exposed for 18 hrs.

1913 nucleotides) of 3' noncoding sequences were due to the alternative polyadenylation signals used. Northern blot analysis estimated the size of PAP mRNAs from benign prostate hyperplasia and canerous prostate to be 3.2Kb. These results indicate that the predominent mRNAs resulted from the usage of the 3' downstream polyadenylation signal.

The PAP cDNA sequence from the newly isolated clone hP303 was found to be identical to that of the 10 exons present in the human PAP gene. This cDNA sequence is also identical to that reported by van Etten et al. (10). The PAP cDNA sequences described by Vihko et al. (11) and Tailor et al.(12) exhibited some nucleotide differences, which might be due to genetic polymorphism and/or sequencing errors. The deduced amino acid sequence of PAP contains a leader peptide

of 32 residues and secretory enzyme of 354 amino acids as described previously (6,10-12).

The amino acid sequence of human PAP was reported (6) to exhibit 50% (177/351) identity to the corresponding sequence of human lysosomal acid phosphatase (LAP). The protein-coding sequence of 423 amino acids from the human LAP gene was shown to be interrupted by 10 introns (13). The positions of eight introns (1 to 6, 8 and 9) present in the human PAP gene were found to be homologous to the sites of corresponding introns from human LAP gene. The seventh intron of PAP gene is located between codon no. 260 and 261 (Gly-Gly at amino acid no. 228-229 of secretory enzyme), while that of the LAP gene is located within codon 257 (equivalent to amino acid 229 of the secretory PAP enzyme). This difference could be due to the difficulty of sequencing the string of six G's in the LAP gene (13). The 10th intron of the LAP gene was located within codon 380, which corresponds to codon 382 of the PAP gene (amino acid 350 of the secretory PAP enzyme). The last (11th) exon of the LAP gene encodes the 43 COOH-terminal amino acids, including a transmembrane segment, which is absent in the PAP gene. The PAP and LAP genes, which are located on human chromosome no. 3 and 11, respectively, evolved from duplication and divergence of an ancestral acid phosphatase gene. It has been suggested that the ancestral exon 11 was lost by deletion in the human PAP gene (13). Alternatively, the exon 11 of the LAP gene could have been added after gene duplication.

Southern blot analysis of ten mammalian genomic DNAs (Fig. 5) indicated multiple EcoRI fragments of human PAP gene. This is consistent with the exon-intron organization of human PAP gene (Fig. 2) elucidated in this investigation. Whereas most of the other mammalian DNAs gave three bands, indicating the presence of at least two EcoRI sites in their PAP genes.

The synthesis of human PAP by epithelial cells of the prostate is induced by androgen (1). The putative promoter region of human PAP gene remains to be characterized and the investigation on the molecular mechanism(s) of androgen

induction might provide important information on the androgen dependency of prostate cancer (14).

ACKNOWLEDGMENTS

We thank Drs. Y. Sharief and J. Mohler for providing the human tissues, Drs. F. Johnson and Y. H. Liu for reading the manuscript and Ms. N. Gore for preparing it.

REFERENCES

1. Yam, L.T. (1974) Am. J. Med. 56, 604-616.

2. Gutman, E.B., Sproul, E.E., and Gutman, A.B. (1936) Am. J. Cancer 28, 485-495.

3. Chu, T.M., Wang, M.C., Lee, C.L., Killian, C.S., and Murphy, G.P. (1982) In Biochemical Markers for Cancer (Chu, T.M., ed.) pp 117-136, Marcel Dekker, New York.

4. Lin, M.F., and Clinton, G.M. (1987) Adv. Protein Phosphatases 4, 199-228.

5. Lin, M.F., Li, S.S.-L., Chu, T. M., and Lee, C.-L. (1990) J. Clin. Lab. Anal. 4, 420-425.

6. Sharief, F.S., Lee, H., Leuderman, M.M., Lundwall, A., Deaven, L.L., Lee, C.-L., and Li, S.S.-L. (1989) Biochem. Biophys. Res. Commun. 160, 79-86.

7. Lee, H., Lee, C.-L., and Li, S.S.-L. (1991) J. Protein Chem. 10, 253-256.

8. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning, a laboratory manual, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

9. Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 3963-3965.

10. Van Etten, R.L., Davidson, R., Stevis, P.E., MacArthur, H., and Moore, D.L. (1991) J. Biol. Chem. 266, 2313-2319.

11. Vihko, P., Virkkunen, P., Henttu, P., Roiko, K., Solin, T., and Huhtala, M.-L. (1988) FEBS Letters 236, 275-281.

12. Tailor, P.G., Govindan, M.V., and Patel, P.C. (1990) Nucleic Acids Res. 18, 4928.

13. Geier, C., Figura, K.v., and Pohlmann, R. (1989) Eur. J. Biochem. 183, 611-616.

14. Chiarodo, A. (1991) Cancer Res. 51, 2498-2505.